Evolution of a CAM anatomy predates the origins of Crassulacean acid metabolism in the Agavoideae (Asparagaceae)

Karolina Heyduk^a, Michael R. McKain^b, Falak Lalani^a, James Leebens-Mack^a.

^aDepartment of Plant Biology, 2502 Miller Plant Sciences, University of Georgia, Athens, GA 30602; ^bDonald Danforth Plant Science Center, 975 N. Warson Rd., St. Louis, MO 63132

Author for correspondence:

Karolina Heyduk

Tel: 1-706-542-1830

Email: heyduk@uga.edu

1 **Abstract**

2 Crassulacean acid metabolism (CAM) is a modified form of photosynthesis that has arisen
3 independently at least 35 times in flowering plants. The occurrence of CAM is often correlated
4 with shifts to arid, semiarid, or epiphytic habits, as well as transitions in leaf morphology (e.g.
5 increased leaf thickness) and anatomy (e.g. increased cell size and packing). We assess shifts
6 between C₃ and CAM photosynthesis in the subfamily Agavoideae (Asparagaceae) through
7 phylogenetic analysis of targeted loci captured from the nuclear and chloroplast genomes of over
8 60 species. Carbon isotope data was used as a proxy for mode of photosynthesis in extant species
9 and ancestral states were estimated on the phylogeny. Ancestral character state mapping suggests
10 three independent origins of CAM in the Agavoideae. CAM species differ from C₃ species in
11 climate space and are found to have thicker leaves with densely packed cells. C₃ ancestors of
12 CAM species show a predisposition toward CAM-like morphology. Leaf characteristics in the
13 ancestral C₃ species may have enabled the repeated evolution of CAM in the Agavoideae
14 subfamily. Anatomical changes, including a tendency toward 3D venation, may have initially
15 arisen in C₃ ancestors in response to aridity as a way to increase leaf succulence for water
16 storage.

17

18 Keywords: CAM; character evolution; leaf anatomy; photosynthesis; phylogenomics; sequence
19 capture

1) Introduction

When RuBisCO (Ribulose biphosphate carboxylase oxygenase) evolved some 3 billion years ago, Earth's atmosphere was high in CO₂ and lacking oxygen. Today, however, the 20% of the atmosphere that is oxygen can be fixed by RuBisCO, engaging the photorespiratory pathway in C₃ plants. Photorespiration is energetically costly, and particularly pronounced under high temperatures or where the ratio of CO₂ to O₂ is depleted in leaves. C₃ plants that are subjected to regular water stress will close their stomata to conserve water and draw down the internal leaf carbon concentration. To circumvent photorespiration due to RuBisCO's affinity for O₂, many photosynthetic organisms have developed carbon concentrating mechanisms (CCMs) to enrich the air around RuBisCO with carbon; it's estimated that half of all carbon fixed is acquired through a CCM (Raven et al., 2008). Crassulacean acid metabolism (CAM) is a complex CCM that has evolved from C₃ ancestors at least 35 times in independent plant lineages (Silvera et al., 2010). CAM plants conduct the majority of their gas exchange at night, when open stomata are subjected to much lower vapor pressure deficits, leading to reduced transpiration rates. Incoming CO₂ is fixed by phosphoenolpyruvate carboxylase (PEPC) and stored as malic acid in the vacuole. In the daytime, the stomata close and malic acid is decarboxylated, resulting in increased concentrations of CO₂ around RuBisCO. The higher concentration of carbon within CAM plant cells results in an increased water use efficiency, and allows CAM plants to grow in some of the most water stressed environments on earth. Phylogenetic analyses suggest that CAM evolved in multiple plant lineages 5-10 Mya (e.g. Good-Avila et al. 2006; Ocampo and Columbus 2010; Bone et al. 2015) and may have allowed species diversification into desert habitats that arose at this time (Axelrod, 1979).

Though CAM is not as phylogenetically clustered as C₄ photosynthesis (Edwards et al., 2010; Sage et al., 2011), a few large lineages do seem to exhibit repeated origins of CAM. High densities of CAM species are found in the Cactaceae, Euphorbiaceae, Crassulaceae in the eudicots, and in the monocot families Bromeliaceae, Orchidaceae, and Asparagaceae (subfamily Agavoideae). CAM has been found in 35 separate lineages (Silvera et al., 2010), but this is likely an underestimate, as broad-scale surveys are often unable to detect intermediate or weak forms of CAM. For example, surveys of CAM plants across broad phylogenetic scales use carbon isotope ratios ($\delta^{13}\text{C}$) as a proxy for mode of photosynthesis: CAM and C₄ plants, via the enzyme PEPC, discriminate less against the heavier ¹³C isotope of carbon than RuBisCO does, resulting

in a ratio between ^{13}C and ^{12}C that is more similar to atmospheric ratios in CAM plants relative to C_3 . CAM plants have $\delta^{13}\text{C}$ ratios between -10‰ and -20‰, while C_3 plants have ratios less than -22‰. Intermediate CAM species, however, act as cryptic CAM plants in that they can have carbon isotope ratio signatures that are within the range of C_3 species (Silvera et al., 2005).

Anatomically, CAM plants from independent lineages have converged on a suite of traits that are directly related to the ability of the CAM pathway to function efficiently. Large cells, and particularly large vacuoles, are required for storing accumulated malic acid, which directly influences the carbon available to RuBisCO during the day. To accommodate large cells, CAM leaves typically exhibit reduced intercellular airspace (IAS), which has the added benefit of limiting diffusion of decarboxylated CO_2 in CAM plant cells during the day (Nelson and Sage, 2008; Nelson et al., 2005; Zambrano et al., 2014). A number of CAM traits represent tradeoffs with the C_3 photosynthetic pathway: for example, a lack of IAS in C_3 plants will limit the ability of CO_2 to diffuse throughout the layers of mesophyll cells. The transition from C_3 photosynthesis to CAM includes the evolution of anatomical modifications that allow for optimal CAM pathway function.

In the C_4 literature, a stepping-stone model of evolution has been proposed in which plants gradually progress from C_3 to C_4 through an intermediate stage known as C_2 (Monson and Rawsthorne, 2000; Sage, 2004). Evolutionary studies of C_3 to C_4 transitions have found that anatomical differences arise before a fully optimized C_4 pathway (Christin et al., 2013, 2011). How CAM lineages evolved from C_3 ancestors is less well described, but a few models have been proposed. Because C_3 plants have the major CAM pathway components functioning in stomatal guard cells, it has been suggested that CAM evolution required shifting these reactions to mesophyll cells (Cockburn, 1981). Alternatively, intermediate or weak CAM species have been implicated in the evolutionary transition from C_3 photosynthesis to full CAM; there is some evidence from the Crassulaceae that suggests CAM cycling or a similar weak form as an evolutionary stepping stone to full constitutive CAM (Teeri, 1982). Sage (2002) suggested that leaf and cell succulence may arise first in a C_3 ancestor, followed by evolution of PEPC function to recapture respired CO_2 , eventually leading to Circadian control and full-fledged CAM function. Succulent anatomy as a precursor for CAM evolution is an attractive hypothesis, especially for plant lineages where succulence could initially alleviate water stress via water storage before being co-opted for CAM function. However, few studies have examined ancestral

characteristics in extant CAM lineages to determine whether ancestral anatomy facilitated CAM evolution. A unified model of C₃ to CAM evolution requires comparative analyses of anatomy and selective forces in closely related C₃ and CAM lineages in order to understand whether certain CAM-like traits appear before the origin of CAM, thus enabling the repeated evolution of a complex trait. Here, we utilize the bimodal karyotype (BK) clade of the Agavoideae (sensu McKain et al. 2012), a subfamily of the Asparagaceae (APG III, Chase et al. 2009), to assess whether CAM anatomical traits arose before or concurrently with the origins of CAM.

The BK clade of the Agavoideae consists of ~600 species and includes the iconic desert genera *Yucca* and *Agave*, in addition to a number of genera inhabiting mesic environments, like *Camassia* and *Hosta*. The subfamily is widespread but largely restricted to the new world (although *Hosta* is an exception, with an eastern Asian distribution (Rocha et al., 2006)). The deserts of North and Central America have been implicated as the center of diversity for several genera within the Agavoideae. To date, analyses of relationships among Agavoideae lineages have been confounded by low and mixed phylogenetic signals (Archibald et al., 2015; Good-Avila et al., 2006; Smith et al., 2008), particularly within genera. The phylogenetic difficulties largely arise from the young age of the group, as the core Agavoideae are estimated to have diverged 20-26 Mya, while *Agave* and *Yucca* represent much more recent diversification events at 8-10 and 9-17 Mya, respectively (Good-Avila et al., 2006; Smith et al., 2008). Phylogenetic resolution of a young lineage can be further confounded by hybridization and incomplete lineage sorting (both of which can lead to gene tree-species tree incongruence (Degnan and Rosenberg, 2009; Maddison, 1997; Rosenberg, 2002); additionally, some species within the Agavoideae are long-lived perennials whose generation time results in slower mutation accumulation.

Research on mode of photosynthesis within the Agavoideae has focused almost exclusively on the physiology and anatomy of *Agave* species (Nobel 1976, 1988, but see Smith et al. 1983; Huxman et al. 1998 for examples in *Yucca*), and the related desert species were assumed to be CAM. We assessed the photosynthetic pathway in over 60 species from 12 genera using carbon isotope ratios and ancestral state estimation on a species tree inferred from analysis of 272 nuclear loci and ~70kb on the plastid genome, sampled by target locus enrichment (Mamanova et al., 2010; Heyduk et al., 2016b). Ancestral state reconstruction suggests multiple origins of CAM and at least one potential reversal from constitutive CAM. In addition, leaf cross sections were examined to assess whether certain CAM promoting traits – such as large cells or

decreased IAS – evolved prior to or concurrently with CAM photosynthesis.

2) Materials and Methods

2.1 - Tissue sampling

For phylogenetic analyses, many of the *Yucca* tissues were collected from natural populations by Olle Pellmyr and collaborators and stored at -80°C for 10-20 years, while tissue for all *Agave* accessions came from the Desert Botanical Garden's live collection (Phoenix, AZ, USA). Additional sources of tissue are described in Supporting Information Table S1. For each species, the source of tissue for carbon isotope measurements is also indicated in Supporting Information Table S1. Most isotope tissue was sourced from the Missouri Botanical Garden Herbarium, the University of Georgia State (UGA) Herbarium, and the Desert Botanical Garden Herbarium. Tissue for anatomical cross sections was harvested from plants growing at the UGA plant biology greenhouses, with samples taken from central portions of fully mature leaves.

2.2 - RNA target enrichment bait design

Following McKain et al. 2012, transcriptome data from *Chlorogalum pomeridianum*, *Chlorophytum rhizopendulum*, *Hesperaloe parviflora*, *Hosta venusta*, *Yucca brevifolia*, and *Yucca filamentosa* were assembled using Trinity r2011-08-20 (Grabherr et al., 2011), with *C. rhizopendulum* acting as the outgroup to the Agavoideae BK clade. Additionally, available EST sequences from *Agave tequilana* (from Simpson et al. 2011) were included in orthogroup clustering and probe design. Transcripts of each species were translated using the RefTrans pipeline (<https://github.com/mrmckain/RefTrans>). Protein models for translations were chosen by BLAST (tblastx) against 10 sequenced angiosperm genomes (http://fgp.bio.psu.edu/tribedb/10_genomes/; Wall et al. 2008). The best hit of each transcript, as determined by e-value and overlapping similarity, was then used for protein alignment using GeneWise v.2.2.0 (Birney et al., 2004). The resulting amino acid and nucleotide coding sequences were clustered into gene families using OrthoMCL (Li et al., 2003) as in McKain et al. (2012). Amino acid sequences in each gene family were aligned in MUSCLE v. 3.8.31 (Edgar, 2004) and the coding sequences were mapped to the protein alignments using PAL2NAL (Suyama et al., 2006). Using rice intron boundaries, each gene family alignment had intron-exon boundaries estimated using the intronFinder.pl script from Citrus Genomics (www.citrusgenome.ucr.edu/usa/ucr/Files.php). Gene alignments were then broken into putative exon sequences and introns were removed for bait design. All sequences within a single exon

alignment had pairwise distance calculated to estimate the percentage of divergence between taxa for a given exon. Sequences were clustered within an exon so that no member of a cluster was more than 10% divergent from any other member thus potentially creating multiple representative alignments for each exon within each gene family alignment. Baits were designed from clustered exon sequences representing 3,709 exons from 2,473 gene families, though only 910 of these were expected to be single-copy gene families.

2.3 - Library construction and enrichment

DNA was isolated from all tissue using a modified CTAB protocol (Doyle, 1987), with an additional sorbitol cleaning step to remove plant secondary compounds (Štorchová et al., 2000). DNA was sheared with a Covaris sonicator (Covaris Inc.) to an average size of 350bp. Libraries were constructed using either single-ended or dual-indexed adapters using a protocol modified from Fisher et al. (2011). Libraries were then used as the template for hybridization with the RNA baits as described in the Myselect protocol (www.mycroarray.com) and by Heyduk et al. (2016b). Briefly, libraries are hybridized to biotinylated RNA baits for 36 hours, with 4-5 libraries per pool. Nuclear and chloroplast baits were added to the same hybridization reaction, at a 4:1 nuclear:chloroplast volume ratio. Nuclear exon baits were added at original concentration, but chloroplast genome baits were diluted to 1/500x prior to hybridization. Libraries were pooled based on phylogenetic distance to minimize preferential hybridization and capture; when possible, species from a single genus were pooled together, and cross-genus pools were avoided. Post-hybridization, pools were assessed for target enrichment using quantitative real time PCR (qRT-PCR) of two nuclear targets and a chloroplast target. Abundance of targeted locus fragments were compared to non-targeted ribosomal ITS copy numbers in the enriched pools, using un-enriched libraries as a control. Enriched pools were then quantified using qRT-PCR and pooled for sequencing. Three separate sequencing runs were carried out, including one preliminary run on an Illumina MiSeq using 150bp PE reads, one run on an Illumina HiSeq 2500 with 100bp PE reads, and a final sequencing run on a HiSeq 2500 rapid run with 150bp PE reads, resulting in a final total of 71 species sequenced.

2.4 - Assembly

Reads were quality trimmed and adapters were removed using FASTX-Toolkit v. 0.0.13.2 (http://hannonlab.cshl.edu/fastx_toolkit/index.html) and FAR v. 2.15 (<https://wiki.gacrc.uga.edu/wiki/FAR>), respectively. Libraries for two samples of *Camassia*

quamash and the sample of *Chlorophytum rhizopendulum* were removed due to low read counts. Trinity (release: 2014-04-13-p1) (Grabherr et al., 2011) was used for assembly as sequence capture data, much like RNA-seq for which Trinity was designed, consists of short fragments with variable coverage. Trinity isoforms were collapsed using CAP3 (Huang and Madan, 1999) with at least 95% identity and a 20bp overlap. Nuclear contigs were then matched against the original exon reference sequences used to design baits with blastn. If multiple assembled contigs hit the same exon, they were excluded from downstream analysis due to the inability to differentiate paralogs from alleles. For chloroplast alignments, cleaned reads were first mapped to a reference sequence from *Yucca filamentosa* (McKain et. al, *unpublished*) using Bowtie2 (Langmead and Salzberg, 2012). Only reads that mapped to the reference were used in subsequent YASRA (Ratan, 2009) assemblies, each of which required a minimum contig coverage of 10x for assembly. Contigs from YASRA were merged end-to-end for each species. Coverage for chloroplast and nuclear assemblies was calculated using Bowtie v.2 (Langmead and Salzberg, 2012) and bedtools using the genomeCoverageBed option (Quinlan and Hall, 2010).

2.5 - Phylogenetic reconstruction

To reconstruct phylogenetic relationships, only target sequences were kept using methods described in Heyduk et al. 2016b, but were modified as follows. Exons were merged together into genes based on their order in the original target sequences. Genes were then aligned using PRANK (Löytynoja and Goldman, 2005) and cleaned with Gblocks (Castresana, 2000), requiring half of the taxa for a given gene to have sequence in a block for it to be kept. To reduce spurious alignments, sequences within each gene alignment were removed if they did not comprise at least 50% of the total alignment length. Genes were only considered for further analyses if they were at least 200bp long, contained at least one parsimony informative single nucleotide polymorphism, and had at least 35 of the 68 taxa in the alignment. Gene trees were estimated independently in RAxML (Stamatakis, 2006) with 500 bootstrap replicates and the GTRGAMMA model of nucleotide substitution.

The RAxML best trees and bootstrap trees were used in a subsequent ASTRAL (Mirarab et al., 2014) analysis with 400 bootstrap replicates. In a few species, multiple individuals were sequenced but were treated independently (versus being forced to represent a single lineage in the species tree). ASTRAL bootstrap replicate trees were imported into Geneious 8.1.6 (Kearse

et al., 2012) to estimate a majority-rule consensus tree using a threshold of 50 bootstrap support. In addition, individual gene alignments were also concatenated into a supermatrix, and a species tree was estimated using RAxML and the GTRGAMMA model as above. Species relationships were estimated in RAxML with GTRGAMMA and 500 bootstrap replicates. ASTRAL implements a coalescence-based approach to species tree estimation, whereas the supermatrix analysis assumes that all genes share the same history and does not take into account deep coalescence that arises from incomplete sorting of ancestral alleles during speciation events. The ASTRAL and supermatrix trees were compared to assess the influence of gene tree discordance in supermatrix analysis. Chloroplast sequences were aligned with MAFFT (Katoh and Standley, 2013) and were cleaned with Gblocks with the same parameters as used in the nuclear data.

To assess the degree of conflict between gene trees and estimated species trees, gene trees were counted for the presence or absence of specific clades that differed in their placement in the concatenated and ASTRAL tree estimates. A clade was counted as present in a gene tree regardless of the bootstrap support in the gene tree, though we consider strong support only those gene trees that have a bootstrap value of 50 or greater. Conflict was assessed using getConflict.pl, available at <https://github.com/kheyduk/Phylogenomics>.

2.6 - Isotopic analysis

Tissue for isotopic analysis was largely collected from either the Desert Botanical Garden herbarium (Phoenix, AZ, USA), the Missouri Botanical Gardens (St. Louis, MO, USA), or from the Georgia State Herbarium (Athens, GA, USA) (for all accession information, see Supporting Information Table S1). Tissue was selected from non-horticultural accessions whenever possible to avoid effects of watering on the plasticity of mode of carbon uptake. Samples were dried at 50°C for two days then hand-ground with mortar and pestle. Isotopic composition was assessed at the UGA Soil Analysis Laboratory against a background of Pee Dee belemnite.

2.7 - Ancestral state reconstruction

Estimation of ancestral states at the nodes of the Agavoideae tree was conducted in diversitree (FitzJohn, 2012) in R 3.2.2 (R Core Team, 2016) using a fully-bifurcating version of the supermatrix tree, which we forced to be ultrametric via the chronos function in ape (Paradis et al., 2004) which uses penalized likelihood (Sanderson, 2002). A number of taxa did not have carbon isotope values associated with them, and the following were pruned from the tree before ancestral state reconstruction: *Agave delamateri*, *Agave palmeri*, *Camassia leichtlinii* spp.

suksdorfii, *Camassia quamash* spp. *maxima*, *Hesperaloe nocturna*, *Hesperaloe funifera*, *Hesperaloe campanulata*, and *Yucca lousisianensis*. For *Beschorneria*, the carbon isotope value of *Beschorneria yuccoides* was used as a representative for the genus. Likewise, for the genera of *Hosta* and *Echeandia*, photosynthetic pathway was estimated from congeners of the species represented in the molecular phylogeny (Supporting Information Table S3). Although diversification rates might vary between C₃ or CAM lineages, simulation work has shown that models that incorporate diversification along with transitions from one character state to another can be inaccurate when the number of taxa is low (BiSSE, Davis et al. 2013). Instead, a Markov model of discrete character evolution was used to first assess whether transitions from C₃ to CAM were equally likely as reversions from CAM to C₃ in the diversitree package in R. As there was no significant improvement in model fit via an ANOVA when allowing for different transition rates (p=0.44), a single rate was used to estimate ancestral states at the nodes of the phylogeny using marginal probabilities.

2.8 - Climate space analysis

GPS coordinates were taken from herbarium records of species used in the phylogeny, excluding records from gardens or those from non-native ranges. For those coordinates, 19 bioclimatic variables were downloaded from the WorldClim database (<http://www.worldclim.org/bioclim>) using the raster package in R (Hijmans and van Etten, 2012) with a spatial resolution of 2.5 minutes of a degree. A principal coordinate analysis using the prcomp function in R was used to assess climate similarities between estimated C₃ and CAM species. Samples of species were designated C₃ or CAM according to the exemplar species with measured isotope ratios in this study; although this may be a generalization for more intermediate species, the majority of species are expected to be fixed for CAM or C₃ (Silvera et al., 2005) and can be represented by a single isotope value.

2.9 - Anatomical evolution

Cross sections from exemplars from the following species were collected from fresh tissue: *Agave schottii*, *Agave palmeri*, *Beschorneria yuccoides*, *Chlorophytum rhizopendulum*, *Hesperaloe funifera*, *Hesperaloe parviflora*, *Hosta ventricosa*, *Manfreda scabra*, *Manfreda virginica*, *Polianthes tuberosa*, *Yucca brevifolia*, and *Yucca angustissima*. Leaves were cut, fixed in formalin, embedded in paraffin, sectioned with a microtome, and stained with Toluidine blue at the UGA Veterinary Histology Laboratory. When available, two separate plants per species

were used for cross sectioning; for every cross section, 2-3 images were captured across the leaf, measured, and then averaged. Cross sections traits were measured as in Heyduk et al. (2016a) and are described here briefly. Leaf thickness, average cell size (adaxial and abaxial), intercellular air space (IAS), and the number of planes of fully developed vasculature were counted and measured in ImageJ (Rashband, n.d.). IAS was measured as a percent of mesophyll area. Student's T-test was used to compare CAM and C₃ species for anatomical traits.

2.10 - Data availability

All raw sequencing data is available through BioProject PRJNA316210 at Genbank's SRA database. Gene alignments, gene trees, and species trees in newick format are available on Dryad ([doi:xxx](#)), along with cross-sectional phenotyping measurements.

3) Results

3.1 - Sequence capture

After read-cleaning, libraries had an average of 7 million reads. See Supporting Information Table S2 for library-specific read mapping rates and coverage. On average, 1232 single-copy exons were captured per library representing an average of 899 genes. After removing alignments that were too short, uninformative, or those that had less than 34 taxa or were missing an outgroup, 272 genes remained. Average coverage per exon from cleaned set of 272 genes was 64X, while the adjacent intron sequences had an average coverage of 63X (for library-specific coverage values, see Supporting Information Table S2). Average coverage over the ~70kb of chloroplast genome assembled was 121X (Supporting Information Table S2). The baits designed were successful across the entire subfamily, including outgroups *Anemarrhena* and *Echeandia*.

3.2 - Phylogenetic inference

Although relationships among genera of the Agavoideae have been fairly well described to date (Archibald et al., 2015; McKain et al., 2012; Smith et al., 2008), species-level relationships have proven to be more difficult to resolve with any confidence, despite interest in both *Yucca* and *Agave* for pollination syndrome evolution (Pellmyr, 2003; Rocha et al., 2006). Using 272 nuclear genes spanning a total of 123,651bp, including both exonic and adjacent non-coding sequences, we were able to resolve fine scale relationships among genera (Fig. 1). Aside from the well-documented paraphyly of *Agave* (e.g. Bogler and Simpson 1996; Bogler et al. 2005; Rocha et al. 2006; Smith et al. 2008), all genera were supported as monophyletic, with the

exception of *Manfreda* and *Hesperaloe* in the chloroplast analyses (Fig. 1c).

Resolution within the genera *Agave* and *Yucca* is not much improved from previous analyses and is likely a result of a rapid radiation of each genus. For *Agave sensu lato* in particular, the polytomy at the backbone of the genus suggests current or recent gene flow among species (Gentry, 1982) as well as retention of ancestral polymorphism due to short branches on the tree (Fig. 1). Even with 272 genes and ~70kb of chloroplast sequence, phylogenetic resolution within this genus is poor, suggesting a hard polytomy due to rapid radiation rather than a bifurcating speciation process. In agreement with previous analyses of plastid and ribosomal markers (Bogler and Simpson, 1996; Bogler et al., 2005; Rocha et al., 2006; Smith et al., 2008), the placement of *Manfreda* and *Polianthes* remains within *Agave*, indicating a recent divergence of these genera from *Agave* ancestors. Relationships among *Yucca* species are more resolved, with the three subgenera typically identified by their fruit types – *Chaenocarpa* (dehiscent fruit), *Sarcocarpa* (fleshy fruit), and *Clistocarpa* (spongy, monospecific containing only *Y. brevifolia*) forming well-supported clades in all three analyses (Fig. 1). The two samples of *Y. brevifolia* form a clade that is sister to the rest of *Yucca* in both nuclear analyses but sister only to *Chaenocarpa* in the plastid tree. Furthermore, the position of *Y. pallida* is unstable regarding its placement either as sister to the other sampled *Chaenocarpa* species (Fig. 1a, 1b) or nested deeply within *Chaenocarpa* (Fig. 1c).

Relationships among major lineages in the Agavoideae were well resolved in the nuclear gene analyses with the exception of the placement of *Hosta*. Whereas the plastome and supermatrix analysis provide strong support for *Hosta* as sister to the rest of the Agavoideae BK clade (Figs. 1b and 1c), the ASTRAL analysis of nuclear genes does not resolve the relationships among *Hosta*, the *Camassia*+*Chlorogalum*+*Hastingsia* clade, and the *Hesperaloe*+*Schoenolirion*+*Hesperoyucca* clade (Fig 1a). Individual gene trees were assessed for their support for the two alternative locations of *Hosta* – either sister to the remainder of the BK clade, or more closely related to *Camassia* and *Hesperaloe* clades. Of the 272 gene trees, *Hosta* as sister to the remainder of the BK clade was found in 153 gene trees (of those, only 19 had bootstrap support >50 for the BK node without *Hosta*). In comparison, only 26 gene trees supported *Hosta* as sister to *Camassia* and *Hesperaloe* clades (17 had bootstrap support >50).

3.3 - Ancestral states and climate space

Carbon isotope values for all species are listed in Supporting Information Table S3; the

cutoff between C₃ and CAM was set to -20‰, though cryptic CAM species can have isotope values well into the C₃ range (Winter and Holtum, 2002; Winter et al., 2015), and extremely water use efficient C₃ species can have isotopic signatures in the C₃-CAM intermediate range (Farquhar and Richards, 1984; Farquhar et al., 1989). Despite these caveats, we chose a hard threshold for C₃ vs. CAM, but species with intermediate values should be screened in the future using more sensitive physiological measures such as gas exchange. Nonetheless the interpretation of carbon isotope values presented here is in line with previous physiological studies on species with intermediate isotope values: for example, *Hesperaloe* (CAM) (Ravetta and McLaughlin, 1993) and *Yucca brevifolia* (C₃) (Smith et al., 1983). Additionally, while photosynthetic pathway of *Hosta* is based on isotope values from only a single accession, previous gas exchange measurements on the *Hosta* species (Zhang et al., 2004) and knowledge of anatomy and preferred habitat solidifies our assumption that the majority, if not all, of *Hosta* are C₃. Ancestral state reconstruction on a fully bifurcating supermatrix phylogeny of the Agavoideae showed high support for three independent origins of CAM and one reversion: one gain at the base of *Hesperaloe*, one at the base of subgenus *Sarcocarpa* in *Yucca*, and a third at the most recent common ancestor of *Agave*, *Manfreda*, and *Polianthes* (Fig. 2). A possible reversion is seen at *Polianthes tuberosa*, though this reversal should be confirmed with physiological measurements to rule out an intermediate form of CAM operating in this species. Using age estimates from previous molecular dating studies in the Agavoideae, the evolution of CAM in all three instances occurs at roughly the same time, about 6-10 Myr ago (Ma), based on the ages of the main clades (6-10 Ma for *Yucca*, 8.7-10 Ma for *Agave*) (Smith et al., 2008). The main transitions to CAM from C₃ are placed within clades that are well-supported in all three species tree analyses (*Beschorneria*+*Agave* s.l., *Yucca*, and *Hesperaloe*+*Schoenolirion*+*Hesperoyucca*). The *Yucca* section that evolved CAM is monophyletic and strongly supported as nested within the C₃ *Yucca* species (Figs. 1 and 2), and is therefore estimated to be an independent origin of CAM. Although the position of *Polianthes* is unresolved within the genus *Agave*, it has strong support for its placement within the genus *sensu lato*, implying a significant change in mode of carbon acquisition from *Agave* (with strongly CAM carbon isotope ratios) to *Polianthes* (with a strongly C₃ isotope value).

A PCA of the 19 climate variables from the WorldClim database show a large degree of separation between C₃ and CAM taxa, with 66% of the total variation explained by the first two

axes (Fig. 3). The first component, which separates C₃ from CAM plants, is largely driven by the minimum temperature of the coldest month (loading score of 0.33), the mean temperature of the coldest quarter (0.33), the annual temperature range (-0.32), and temperature seasonality (standard deviation of temperatures) (-0.31). The second axis largely separates taxa by the maximum temperature in the warmest month (-0.38), the mean temperature of the warmest quarter (-0.35), precipitation of the driest quarter (0.33), and precipitation of the driest month (0.32). Temperature, particularly that of the coldest and warmest times of the year, impacts where CAM and C₃ species are distributed, but the large degree in overlap between C₃ and CAM indicates a range of seasonal variation in temperature that permits the co-occurrence of both C₃ and CAM species.

3.4 - Anatomical evolution

Average leaf thickness and average cell area (including both abaxial and adaxial mesophyll cells) were significantly different between C₃ and CAM plants (Table 1). In addition, CAM plants had a larger number of planes of vascular bundles than C₃ plants, although by specifically avoiding the large central mid-rib to minimize heterogeneity in leaf measurements, these results may be biased (unmeasured cross sections from thicker midrib regions are shown in Fig. 4). A number of C₃ species, including *B. yuccoides* and *Y. angustissima*, had a propensity for succulence, albeit restricted to the center of the leaf (Fig. 4). Further, although CAM plants have a greater degree of 3D venation, when assessed together both C₃ and CAM plants in the Agavoideae have significantly more planes of vascular bundles than 1 (one sample T-test, df=37, p=0). Although the degree of intercellular airspace was significantly different between C₃ and CAM species (Table 1), a number of C₃ genera, including *Polianthes* and *Beschorneria*, had values more similar to their CAM relatives in *Agave* (Table 1). We initially included *Chlorophytum rhizopendulum* for leaf anatomical analysis because it was represented in our sequencing efforts, but the library did not have sufficient read coverage generated for phylogenetic use; previous work has placed *Chlorophytum* sister to *Echeandia* (Seberg et al., 2012), and a carbon isotope value of -28.53‰ suggests strong C₃ (Supporting Information Table S3).

4) Discussion

4.1 - Origins and loss of CAM

Historically, phylogenetic analyses of the Agavoideae have been able to identify major

clades, but relationships among these clades has been difficult to resolve (Archibald et al., 2015; Bogler and Simpson, 1996; Bogler et al., 2005; Rocha et al., 2006; Smith et al., 2008). This work, which comprises 272 single copy nuclear genes and an additional ~70kb of chloroplast genome sequence, found strong support for five major clades: *Beschorneria*+*Agave* s.l. , *Yucca*, *Camassia*+*Chlorogalum*+*Hastingia*, *Hesperaloe*+*Schoenolirion*+*Hesperoyucca*, and *Hosta*. This study included comprehensive sampling of the Agavoideae BK clade and many more loci than had been analyzed in previous studies, and was able to resolve a bifurcating set of relationships between the five major clades on the backbone of the Agavoideae phylogeny.

The evolution of CAM within the Agavoideae is of particular interest due to the high diversity of habitats that species in this subfamily inhabit. Species within *Camassia* and *Hosta* are found in mesic regions, including wet meadows for *Camassia* and shady forest understories for species of *Hosta*. In addition, although the large genera of *Agave* and *Yucca* have a center of diversity in the southwestern deserts of North America, a number of species – particularly those nearer the base of the tree – have distributions outside of the iconic desert range. These more anciently derived species’ ranges may suggest that ancestrally the Agavoideae was composed of non-desert dwellers and established lineages migrated into more arid regions after an early radiation within the group. A movement into arid regions would require that those desert regions were in place already, and that species that moved there had an ability to grow in arguably some of the harshest conditions on the planet. It is equally possible that species adapted in place as aridification occurred. In line with hypotheses regarding habitat and CAM evolution being closely linked, niche characterization of Eulophiinae orchids showed that the origin of CAM allowed for colonization of drier habitats (Bone et al., 2015). Deserts in North America stabilized in climate and vegetation between 10 and 2 Ma (Moore and Jansen, 2006; Waters and Haynes, 2001); this date is in line with the estimated age of the origins of CAM (Figure 2, Smith et al., 2008). Further taxon sampling could allow for better estimation of diversification rates in CAM and C₃ lineages to assess whether the ability for CAM species to move into novel habitats increased biodiversity of these lineages.

The degree of overlap in the climate between CAM and C₃ species (Fig. 3) is driven largely by the distribution of *Yucca*, wherein species from separate subgenera are known to be sympatric (Althoff et al., 2012). The overlap in range between C₃ and CAM *Yucca* species suggests that whereas CAM may be an adaptation to arid environmental conditions, some C₃

Yucca growing under arid conditions may not have the molecular building blocks for the evolution of CAM, and/or may have evolved alternative physiological adaptations to extreme water stress. In either case, there is currently no data that speaks to why members of *Chaenocarpa* did not subsequently evolve CAM photosynthesis. C_3 *Yuccas* are by no means marginal species in the deserts of North America; for example, the C_3 species *Yucca elata* can be a dominant component of Sonoran desert plant communities.

The possible reversion to C_3 or intermediate physiology from a CAM ancestor in *Polianthes* is rare case of transition away from the derived CAM state (but see Crayn et al. 2004 for an example in the Bromeliaceae). Given the high-level resolution of the WorldClim database, *Polianthes* is more similar to CAM than other C_3 species in climate space, but generally grows in local habitats that are less arid than most *Agave* habitats. *Polianthes* species are native to Mexico and grow largely in pine-oak forests or grassland habitats (Solano and Feria, 2007) – certainly dry and hot habitats, but not quite as extreme as habitats occupied by other *Agave* s.l. species (including *Polianthes* and *Manfreda*). Species within the sister group, *Manfreda*, occupy dry sandy soils or rock outcrops in Mexico and the South Eastern U.S. (Verhoek, 1975). Whether the shift in habitat preceded the potential loss of constitutive CAM or vice versa remains unknown, however *Polianthes tuberosa* retains similarities to its CAM relatives anatomically, with %IAS levels nearly as low as those in *Agave* species (Fig. 4, Table 1). Sampling additional species of this genus would address whether all species in *Polianthes* show a similar change in carbon isotope values and whether they all share the same loss of ancestral CAM anatomy. Further physiological analyses, such as gas exchange measurements and leaf acid measurements, will clarify whether *Polianthes* is a true loss or a modification of its ancestral CAM pathway.

Finally, it is worthwhile to remember that C_3 and CAM are at two ends of a spectrum, with intermediate modes of photosynthesis in between (Winter et al., 2015). Intermediate C_3 -CAM species are harder to describe through carbon isotopes alone. For example, species that have C_3 carbon isotopes but may be performing CAM photosynthesis at low levels – known as cryptic CAM species (Silvera et al., 2005) – require more detailed physiological analyses beyond carbon isotopes. Variation in carbon isotopes within a species was also not addressed in this study; while most species can be assigned to either C_3 or CAM consistently with a single accessions, others can show variation in carbon isotope values (e.g., *Puya chilensis* (Crayn et al., 2015; Quezada et al., 2014)) making designation of photosynthetic pathway more challenging.

Although use of carbon isotopes paints a broad picture of C₃-CAM evolution, it nevertheless is a useful method of surveying large groups of taxa and can direct future research toward intermediate or potentially cryptic lineages.

4.2 - Evolution of CAM leaf anatomy

Although differences between CAM and C₃ leaves exist in the Agavoideae, particularly thickness and cell size, the number of similarities between CAM and C₃ species suggests this lineage as a whole developed a CAM anatomy prior to the evolution of the CAM pathway. In the Agavoideae BK clade, all taxa sampled for cross sectional anatomy had a degree of 3D venation in that vascular bundles were aligned on more than a single plane, with notable exceptions of *Hosta*, which is likely sister to the Agavoideae BK clade (Figs. 1 and 4) and *Chlorophytum*, which is unplaced in this phylogenetic assessment but is thought to be sister to the outgroup genus *Echeandia* (Seberg et al., 2012). The propensity toward 3D venation was irrespective of photosynthetic pathway. The trend is particularly striking in C₃ *Yuccas*, where central portions of leaves in *Y. angustissima* and *Y. brevifolia* (Fig. 4) are as thick as *Agave* leaves, and in C₃ *Beschorneria*, which is the sister genus to the CAM *Agave* and shows a similar thickening of the leaf and 3D venation restricted to the midrib area. In both of these cases, the predisposition toward 3D venation potentially allowed for an increase in leaf thickness and cell succulence. Large cells and 3D venation were likely selected for as mechanisms to store water initially, as *Beschorneria*, *Y. angustissima*, and *Y. brevifolia* inhabit the deserts of the southwestern U.S. and Mexico. The phylogenetic positions of *Beschorneria* as sister to *Agave* s.l., and *Y. brevifolia* and *Yucca* sect. *Chaenocarpa* as sequential sister lineages to *Yucca* sect. *Sarcocarpa* imply that the last common ancestor of *Yucca* and *Agave* was C₃ with CAM-like leaf anatomy. CAM-like traits may have facilitated the evolution of in CAM in both *Yucca* sect. *Sarcocarpa* and *Agave* lineages without much in the way of structural change in the leaf – large cells were already in place and could be converted to store malic acid nocturnally. The shared anatomical tendencies again raise the question why all species of the Agavoideae are not employing CAM, especially C₃ *Yuccas* that have both anatomical characteristics and the environmental pressure thought to be required for CAM evolution. Anatomical enabling was shown to be similarly important in C₄ grass lineages (Christin et al., 2013), but is suspected to be the result of random changes in the history of grass clades, rather than from any environmental pressures. In other C₃-CAM transitions, C₃ species closely related to CAM lineages have also been shown to have CAM-like

anatomy (IAS in *Zygopetalum intermedium* in the Orchidaceae, Nelson et al., 2005), but additional anatomical studies in a phylogenetic framework are necessary in CAM lineages to elucidate whether anatomical predisposition is conserved in independent origins of CAM.

Additional sampling for cross sectional anatomy of leaves may elucidate patterns that are currently unclear in the Agavoideae. Phenotypically, there is a large jump to thick, large-celled leaves from *Hosta* to *Hesperaloe* (Fig. 4), but that comparison is missing a number of species that may help describe that transition more carefully (Fig. 1). Sampling *Hesperocallis*, *Shoenolirion*, and *Camassia* leaves will better shape our understanding of C₃ to CAM anatomical transitions at the base of the Agavoideae phylogeny. Nonetheless, the possibility that the last common ancestor of all extant CAM Agavoideae lineages was predisposed to the evolution of CAM may implicate ancient polyploidy (McKain et al., 2012) as a contributing factor in the evolution of a novel trait. Polyploidy has been associated with the origin of innovations in other groups including the origin of nodulation in legumes (Cannon et al., 2015) and the origin of the flower (*Amborella* Genome Project, 2013; Jiao et al., 2011), but this speculative hypothesis remains to be tested for CAM evolution.

4.3 - Conclusions

Despite difficulties in resolving the phylogeny of the subfamily Agavoideae, ancestral state reconstruction shows three independent origins of CAM in the group. These origins are associated with a shift in climate space toward warmer, drier habitats; in *Yucca*, the shift is less clear, as ranges between the C₃ and CAM clades overlap, and the driver for CAM evolution in one clade but not the other continues to be an area of interest. Assessment of leaf characters shows a propensity for the subfamily to have thick leaves enabled by 3D venation. It's possible that veins arranged in multiple planes allowed for larger cells for water storage in dry habitats. Larger cells and thicker leaves facilitated the transition from C₃ photosynthesis to CAM, as large cells are required for efficient storage of malic acid. The early evolution of key anatomical traits for CAM in the Agavoideae likely assisted in the repeated evolution of CAM in three separate lineages, occurring at nearly the same time, and allowed for shifts in habitat occupation. Further work on whether diversification rates vary between C₃ and CAM lineages will give insight into how the evolution of CAM might be promoting biodiversity in the Agavoideae.

Acknowledgements

Many thanks to Joni Ward of the Desert Botanical Garden (Phoenix, AZ, USA) for assistance

with tissue collection, Wendy Hogdson from the Desert Botanical Garden for help sampling herbarium sheets, and to Jeremy Ray and Nia Burrell for help with tissue grinding for isotopic analysis. Susan Kephart generously provided *Camassia* and *Hastingsia* tissue for both DNA and isotopic analysis. Funding was provided by NSF grant DEB-1442199 (to J.L.-M.), NSF DDIG DEB-1010905 (to M.M.), the American Society of Plant Taxonomists Graduate Student Research Grant (to K.H.), the University of Georgia Graduate School IIRG (to K.H.), and the UGA Department of Plant Biology Palfrey Award (to K.H.).

References

- Althoff, D.M., Segraves, K.A., Smith, C.I., Leebens-Mack, J., Pellmyr, O., 2012. Geographic isolation trumps coevolution as a driver of yucca and yucca moth diversification. *Mol. Phylogenet. Evol.* 62, 898–906. doi:10.1016/j.ympev.2011.11.024
- Amborella* Genome Project, 2013. The Amborella genome and the evolution of flowering plants. *Science* 342, 1241089. doi:10.1126/science.1241089
- Archibald, J.K., Kephart, S.R., Theiss, K.E., Petrosky, A.L., Culley, T.M., 2015. Multilocus phylogenetic inference in subfamily Chlorogaloideae and related genera of Agavaceae - informing questions in taxonomy at multiple ranks. *Mol. Phylogenet. Evol.* 84, 266–83. doi:10.1016/j.ympev.2014.12.014
- Axelrod, D.I., 1979. Age and origin of the Sonoran Desert vegetation. *Occas. Pap. Calif. Acad. Sci.* 132.
- Birney, E., Clamp, M., Durbin, R., 2004. GeneWise and Genomewise. *Genome Res.* 14, 988–95. doi:10.1101/gr.1865504
- Bogler, D.J., Pires, C.J., Francisco-Ortega, J., 2005. Phylogeny of the Agavaceae based on *ndhF*, *rbcL*, and its sequences: implications of molecular data for classification. *Aliso* 22, 311–326.
- Bogler, D.J., Simpson, B.B., 1996. Phylogeny of the Agavaceae based on ITS rDNA sequence variation. *Am. J. Bot.* 83, 1225–1235.
- Bone, R.E., Smith, J.A.C., Arrigo, N., Buerki, S., 2015. A macro-ecological perspective on crassulacean acid metabolism (CAM) photosynthesis evolution in Afro-Madagascan drylands: Eulophiinae orchids as a case study. *New Phytol.* doi:10.1111/nph.13572
- Cannon, S.B., McKain, M.R., Harkess, A., Nelson, M.N., Dash, S., Deyholos, M.K., Peng, Y., Joyce, B., Stewart, C.N., Rolf, M., Kutchan, T., Tan, X., Chen, C., Zhang, Y., Carpenter, E., Wong, G.K.-S., Doyle, J.J., Leebens-Mack, J., 2015. Multiple polyploidy events in the early

547 radiation of nodulating and nonnodulating legumes. *Mol. Biol. Evol.* 32, 193–210.
548 doi:10.1093/molbev/msu296

549 Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in
550 phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–52.

551 Chase, M.W., Reveal, J.L., Fay, M.F., 2009. A subfamilial classification for the expanded
552 asparagalean families Amaryllidaceae, Asparagaceae and Xanthorrhoeaceae. *Bot. J. Linn.*
553 *Soc.* 161, 132–136. doi:10.1111/j.1095-8339.2009.00999.x

554 Christin, P.-A., Osborne, C.P., Chatelet, D.S., Columbus, J.T., Besnard, G., Hodkinson, T.R.,
555 Garrison, L.M., Vorontsova, M.S., Edwards, E.J., 2013. Anatomical enablers and the
556 evolution of C₄ photosynthesis in grasses. *Proc. Natl. Acad. Sci. U. S. A.* 110, 1381–6.
557 doi:10.1073/pnas.1216777110

558 Christin, P.-A., Sage, T.L., Edwards, E.J., Ogburn, R.M., Khoshravesh, R., Sage, R.F., 2011.
559 Complex evolutionary transitions and the significance of C(3)-C(4) intermediate forms of
560 photosynthesis in Molluginaceae. *Evolution* 65, 643–60. doi:10.1111/j.1558-
561 5646.2010.01168.x

562 Cockburn, W., 1981. The evolutionary relationship between stomatal mechanism, Crassulacean
563 acid metabolism and C₄ photosynthesis. *Plant, Cell Environ.* 4, 417–418. doi:10.1111/1365-
564 3040.ep11604653

565 Crayn, D.M., Winter, K., Smith, J.A.C., 2004. Multiple origins of Crassulacean acid metabolism
566 and the epiphytic habit in the neotropical family Bromeliaceae. *Proc. Natl. Acad. Sci. U. S.*
567 *A.* 101, 3703–8. doi:10.1073/pnas.0400366101

568 Crayn, D.M., Winter, K., Schulte, K., Smith, J.A.C., 2015. Photosynthetic pathways in
569 Bromeliaceae: phylogenetic and ecological significance of CAM and C₃ based on carbon
570 isotope ratios for 1893 species. *Bot. J. Linn. Soc.* 178, 169–221.

571 Davis, M.P., Midford, P.E., Maddison, W., 2013. Exploring power and parameter estimation of
572 the BiSSE method for analyzing species diversification. *BMC Evol. Biol.* 13, 38.
573 doi:10.1186/1471-2148-13-38

574 Degnan, J., Rosenberg, N., 2009. Gene tree discordance, phylogenetic inference, and the
575 multispecies coalescent. *Trends in Ecol. & Evol.* 24(6), 332–340.

576 Doyle, J., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue.
577 *Phytochem Bull* 19, 11–15.

578 Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high
579 throughput. *Nucleic Acids Res.* 32, 1792–7. doi:10.1093/nar/gkh340

580 Edwards, E.J., Osborne, C.P., Strömberg, C.A.E., Smith, S.A., Bond, W.J., Christin, P.-A.,
581 Cousins, A.B., Duvall, M.R., Fox, D.L., Freckleton, R.P., Ghannoum, O., Hartwell, J.,
582 Huang, Y., Janis, C.M., Keeley, J.E., Kellogg, E.A., Knapp, A.K., Leakey, A.D.B., Nelson,
583 D.M., Saarela, J.M., Sage, R.F., Sala, O.E., Salamin, N., Still, C.J., Tipple, B., 2010. The
584 origins of C₄ grasslands: integrating evolutionary and ecosystem science. *Science* (80-.).
585 328, 587–91. doi:10.1126/science.1177216

586 Farquhar, G., Richards, R., 1984. Isotopic composition of plant carbon correlates with water-use
587 efficiency of wheat genotypes. *Aust. J. of Plant. Phys.* 11(6), 539.

588 Farquhar, G., Ehleringer, J., Hubick, K., 1989. Carbon isotope discrimination and
589 photosynthesis. *Ann. Rev. of Plant Phys. and Plant Molec. Bio* 40(1), 503-537.

590 Fisher, S., Barry, A., Abreu, J., Minie, B., Nolan, J., Delorey, T.M., Young, G., Fennell, T.J.,
591 Allen, A., Ambrogio, L., Berlin, A.M., Blumenstiel, B., Cibulskis, K., Friedrich, D.,
592 Johnson, R., Juhn, F., Reilly, B., Shammass, R., Stalker, J., Sykes, S.M., Thompson, J.,
593 Walsh, J., Zimmer, A., Zwirko, Z., Gabriel, S., Nicol, R., Nusbaum, C., 2011. A scalable,
594 fully automated process for construction of sequence-ready human exome targeted capture
595 libraries. *Genome Biol.* 12, R1. doi:10.1186/gb-2011-12-1-r1

596 FitzJohn, R.G., 2012. Diversitree : comparative phylogenetic analyses of diversification in R.
597 *Methods Ecol. Evol.* 3, 1084–1092. doi:10.1111/j.2041-210X.2012.00234.x

598 Gentry, H.S., 1982. *Agaves of continental North America*. The University of Arizona Press,
599 Tuscon, AZ.

600 Good-Avila, S. V, Souza, V., Gaut, B.S., Eguiarte, L.E., 2006. Timing and rate of speciation in
601 *Agave* (Agavaceae). *Proc. Natl. Acad. Sci. U. S. A.* 103, 9124–9.
602 doi:10.1073/pnas.0603312103

603 Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D. a, Amit, I., Adiconis, X.,
604 Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A.,
605 Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N.,
606 Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a
607 reference genome. *Nat. Biotechnol.* 29, 644–52. doi:10.1038/nbt.1883

608 Heyduk, K., Burrell, N., Lalani, F., Leebens-Mack, J., 2016a. Gas exchange and leaf anatomy of
609 a C₃-CAM hybrid, *Yucca gloriosa* (Asparagaceae). *J. Exp. Bot.* 67, 1369–1379.
610 doi:10.1093/jxb/erv536

611 Heyduk, K., Trapnell, D.W., Barrett, C.F., Leebens-Mack, J., 2016b. Phylogenomic analyses of
612 species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. *Biol.*
613 *J. Linn. Soc.* 117.

614 Hijmans, R.J., van Etten, J., 2012. raster: geographic analysis and modeling with raster data. R
615 package version 3.2.2. <http://CRAN.R-project.org/package=raster>

616 Huang, X., Madan, a, 1999. CAP3: A DNA sequence assembly program. *Genome Res.* 9, 868–
617 77.

618 Huxman, T.E., Hamerlynck, E.P., Loik, M.E., Smith, S.D., 1998. Gas exchange and chlorophyll
619 fluorescence responses of three south-western *Yucca* species to elevated CO₂ and high
620 temperature. *Plant, Cell Environ.* 21, 1275–1283. doi:10.1046/j.1365-3040.1998.00396.x

621 Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E.,
622 Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., Soltis, D.E., Clifton, S.W., Schlarbaum, S.E.,
623 Schuster, S.C., Ma, H., Leebens-Mack, J., dePamphilis, C.W., 2011. Ancestral polyploidy
624 in seed plants and angiosperms. *Nature* 473, 97–100. doi:10.1038/nature09916

625 Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7:
626 Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780.
627 doi:10.1093/molbev/mst010

628 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S.,
629 Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., Drummond, A.,
630 2012. Geneious Basic: an integrated and extendable desktop software platform for the
631 organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649.

632 Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9,
633 357–9. doi:10.1038/nmeth.1923

634 Li, L., Stoeckert, C.J., Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for
635 eukaryotic genomes. *Genome Res.* 13, 2178–89. doi:10.1101/gr.1224503

636 Löytynoja, A., Goldman, N., 2005. An algorithm for progressive multiple alignment of
637 sequences with insertions. *Proc. Natl. Acad. Sci. U. S. A.* 102, 10557–62.
638 doi:10.1073/pnas.0409137102

639 Maddison, W., 1997. Gene trees in species trees. *Syst. Biol.* 46(3), 523–536.

640 Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E.,
641 Shendure, J., Turner, D.J., 2010. Target-enrichment strategies for next- generation
642 sequencing. *Nature Methods* 7(2), 111–118. doi: 10.1038/nmeth.1419

643 McKain, M.R., Wickett, N., Zhang, Y., Ayyampalayam, S., McCombie, W.R., Chase, M.W.,
644 Pires, J.C., dePamphilis, C.W., Leebens-Mack, J., 2012. Phylogenomic analysis of
645 transcriptome data elucidates co-occurrence of a paleopolyploid event and the origin of
646 bimodal karyotypes in Agavoideae (Asparagaceae). *Am. J. Bot.* 99, 397–406.
647 doi:10.3732/ajb.1100537

648 Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T., 2014.
649 ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* 30, i541–
650 8. doi:10.1093/bioinformatics/btu462

651 Monson, R.K., Rawsthorne, S., 2000. Carbon dioxide assimilation in C3-C4 intermediate plants,
 652 in: Leegood, R., Sharkey, T., von Caemmerer, S. (Eds.), *Photosynthesis: Physiology and*
 653 *Metabolism*. Kluwer Academic Publishers, pp. 533–550.

654 Moore, M.J., Jansen, R.K., 2006. Molecular evidence for the age, origin, and evolutionary
 655 history of the American desert plant genus *Tiquilia* (Boraginaceae). *Mol. Phylogenet. Evol.*
 656 39, 668–687. doi:10.1016/j.ympev.2006.01.020

657 Nelson, E.A., Sage, R.F., 2008. Functional constraints of CAM leaf anatomy: tight cell packing
 658 is associated with increased CAM function across a gradient of CAM expression. *J. Exp.*
 659 *Bot.* 59, 1841–50. doi:10.1093/jxb/erm346

660 Nelson, E.A., Sage, T.L., Sage, R.F., 2005. Functional leaf anatomy of plants with Crassulacean
 661 acid metabolism. *Funct. Plant Biol.* 32, 409. doi:10.1071/FP04195

662 Nobel, P.S., 1988. *Environmental Biology of Agaves and Cacti*. Cambridge University Press.

663 Nobel, P.S., 1976. Water relations and photosynthesis of a desert CAM plant, *Agave deserti*.
 664 *Plant Physiol.* 58, 576–582. doi:10.1104/pp.58.4.576

665 Ocampo, G., Columbus, J.T., 2010. Molecular phylogenetics of suborder Cactineae
 666 (Caryophyllales), including insights into photosynthetic diversification and historical
 667 biogeography. *Am. J. Bot.* 97, 1827–47. doi:10.3732/ajb.1000227

668 Paradis, E., Claud, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R
 669 language. *Bioinformatics* 20: 289-290.

670 Pellmyr, O., 2003. *Yucca*, *Yucca* moths, and coevolution: a review. *Ann. Missouri Bot. Gard.* 90,
 671 35–55.

672 Quezada, I.M., Zotz, G., Gianoli, E., 2014. Latitudinal variation in the degree of Crassulacean
 673 acid metabolism in *Puya chilensis*. *Plant Biol.*, 16(4), 848-852.

674 Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic
 675 features. *Bioinformatics* 26, 841–2. doi:10.1093/bioinformatics/btq033

676 R Core Team, 2016. *R: A Language and Environment for Statistical Computing*. R Foundation
 677 for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.

678 Rashband, W.S., 1997-2016. ImageJ. U.S. National Institutes of Health, Bethesda, MD, USA.
 679 <http://imagej.nih.gov/ij/>.

680 Ratan, A., 2009. *Assembly algorithms for next-generation sequence data*. Pennsylvania State
 681 University.

682 Raven, J.A., Cockell, C.S., De La Rocha, C.L., 2008. The evolution of inorganic carbon
683 concentrating mechanisms in photosynthesis. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 363,
684 2641–50. doi:10.1098/rstb.2008.0020

685 Ravetta, D.A., McLaughlin, S.P., 1993. Photosynthetic Pathways of *Hesperaloe funifera* and *H.*
686 *nocturna* (Agavaceae): Novel Sources of Specialty Fibers. *Am. J. Bot.* 80, 524.
687 doi:10.2307/2445367

688 Rocha, M., Good-Avila, S. V, Molina-Freaner, F., Arita, H.T., Castillo, A., Garcia-Mendoza, A.,
689 Silva-Montellano, A., Gaut, B.S., Souza, V., Eguiarte, L.E., 2006. Pollination biology and
690 adaptive radiation of Agavaceae, with special emphasis on the genus *Agave*. *Aliso* 22, 329–
691 344.

692 Rosenberg, N., 2002. The probability of topological concordance of gene trees and species trees.
693 *Theo. Pop. Biol.* 61(2), 225-47.

694 Sage, R.F., 2004. The evolution of C₄ photosynthesis. *New Phytol.* 161, 341–370.
695 doi:10.1111/j.1469-8137.2004.00974.x

696 Sage, R.F., 2002. Are Crassulacean acid metabolism and C₄ photosynthesis incompatible? *Funct.*
697 *Plant Biol.* 29, 775–785. doi:10.1071/PP01217

698 Sage, R.F., Christin, P.-A., Edwards, E.J., 2011. The C(4) plant lineages of planet Earth. *J. Exp.*
699 *Bot.* 62, 3155–69. doi:10.1093/jxb/err048

700 Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: a
701 penalized likelihood approach. *Mol. Biol. Evol.* 19(1):101-109.

702 Seberg, O., Petersen, G., Davis, J.I., Pires, J.C., Stevenson, D.W., Chase, M.W., Fay, M.F.,
703 Devey, D.S., Jørgensen, T., Sytsma, K.J., Pillon, Y., 2012. Phylogeny of the Asparagales
704 based on three plastid and two mitochondrial genes. *Am. J. Bot.* 99, 875–89.
705 doi:10.3732/ajb.1100468

706 Silvera, K., Neubig, K.M., Whitten, W.M., Williams, N.H., Winter, K., Cushman, J.C., 2010.
707 Evolution along the Crassulacean acid metabolism continuum. *Funct. Plant Biol.* 37, 995.
708 doi:10.1071/FP10084

709 Silvera, K., Santiago, L.S., Winter, K., 2005. Distribution of crassulacean acid metabolism in
710 orchids of Panama: evidence of selection for weak and strong modes. *Funct. Plant Biol.* 32,
711 397. doi:10.1071/FP04179

712 Simpson, J., Martinez Hernandez, A., Jazmin Abraham Juarez, M., Delgado Sandoval, S.,
713 Sanchez Villarreal, A., Cortes Romero, C., 2011. Genomic resources and transcriptome
714 mining in *Agave tequilana*. *GCB Bioenergy* 3, 25–36. doi:10.1111/j.1757-
715 1707.2010.01079.x

716 Smith, C.I., Pellmyr, O., Althoff, D.M., Balcázar-Lara, M., Leebens-Mack, J., Segraves, K.A.,
717 2008. Pattern and timing of diversification in *Yucca* (Agavaceae): specialized pollination
718 does not escalate rates of diversification. *Proc. Biol. Sci.* 275, 249–58.
719 doi:10.1098/rspb.2007.1405

720 Smith, S.D., Hartsock, T.L., Nobel, P.S., 1983. Ecophysiology of *Yucca brevifolia*, an
721 arborescent monocot of the Mojave desert. *Oecologia* 60, 10–17. doi:10.1007/BF00379313

722 Solano, E., Ferial, P.T., 2007. Ecological niche modeling and geographic distribution of the genus
723 *Polianthes* L. (Agavaceae) in Mexico: using niche modeling to improve assessments of risk
724 status, in: Hawksworth, D.L., Bull, A.T. (Eds.), *Plant Conservation and Biodiversity*.
725 Springer, pp. 311–326.

726 Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with
727 thousands of taxa and mixed models. *Bioinformatics* 22, 2688–90.
728 doi:10.1093/bioinformatics/btl446

729 Štorchová, H., Hrdličková, R., Chrtěk, J., Tetera, M., Fitze, D., Fehrer, J., 2000. An improved
730 method of DNA isolation from plants collected in the field and conserved in saturated
731 NaCl/CTAB solution. *Taxon* 49, 79–84.

732 Suyama, M., Torrents, D., Bork, P., 2006. PAL2NAL: robust conversion of protein sequence
733 alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34, W609–12.
734 doi:10.1093/nar/gkl315

735 Teeri, J., 1982. Photosynthetic Variation in the Crassulaceae, in: Ting, I.P., Gibbs, M. (Eds.),
736 *Crassulacean acid metabolism*. American Society of Plant Physiologists, Rockville, MD,
737 pp. 244–259.

738 Verhoek, W.S. 1975. A study of the tribe Poliantheae (including *Manfreda*) and revisions of
739 *Manfreda* and *Prochnyanthes* (Agavaceae). PhD thesis. Cornell University.

740 Wall, P.K., Leebens-Mack, J., Müller, K.F., Field, D., Altman, N.S., dePamphilis, C.W., 2008.
741 PlantTribes: a gene and gene family resource for comparative genomics in plants. *Nucleic*
742 *Acids Res.* 36, D970–6. doi:10.1093/nar/gkm972

743 Waters, M.R., Haynes, C.V., 2001. Late Quaternary arroyo formation and climate change in the
744 American Southwest. *Geology* 29, 399–402. doi:10.1130/0091-7613(2001)029

745 Winter, K., Holtum, J.A.M., 2002. How closely do the delta 13C values of Crassulacean acid
746 metabolism plants reflect the proportion of CO₂ fixed during day and night? *Plant Physiol.*
747 129, 1843–1851. doi:10.1104/pp.002915

748 Winter, K., Holtum, J.A.M., Smith, J.A.C., 2015. Crassulacean acid metabolism: a continuous or
749 discrete trait? *New Phytol.* doi:10.1111/nph.13446

750 Zambrano, V.A.B., Lawson, T., Olmos, E., Fernández-García, N., Borland, A.M., 2014. Leaf
751 anatomical traits which accommodate the facultative engagement of crassulacean acid
752 metabolism in tropical trees of the genus *Clusia*. J. Exp. Bot. 65, 3513–23.
753 doi:10.1093/jxb/eru022

754 Zhang, J., Shi, A., Shi, A.P., Zhang, Q.X., 2004. Photosynthetic responses of four *Hosta*
755 cultivars to shade treatments. Photosynthetica, 42(2), 213-218. doi:
756 10.1023/B:PHOT.00000040592.10133.ee

757

Table 1 - Average values (and standard errors) for cross sectional anatomical traits in the Agavoideae.

Species	CAM/C ₃	n ¹	Avg. leaf thickness (um)	Avg. cell area (um)	%IAS	Planes
<i>Agave palmeri</i>	CAM	3	n.a. ²	2888.46±336.74	11.92±3.22	n.a.
<i>Agave schottii</i>	CAM	3	1315.43±370.63	2365.22±188.95	7.53±1.88	1.33
<i>Beschorneria yuccoides</i>	C ₃	3	996.74±262.52	819.35±152.28	11.97±3.99	2.33
<i>Beschorneria yuccoides</i>	C ₃	3	488.53±82.30	853.03±59.86	12.13±2.74	1.67
<i>Chlorophytum rhizopendulum</i>	C ₃	3	228.62±26.32	308.46±10.05	33.07±2.53	1.00
<i>Chlorophytum rhizopendulum</i>	C ₃	2	235.81±2.36	293.55±1.29	33.91±2.01	1.00
<i>Hesperaloe funifera</i>	CAM	3	1282.35±14152	846.19±72.56	4.26±0.23	3.33
<i>Hesperaloe parviflora</i>	CAM	2	1741.93±n.a.	913.64±145.65	7.56±n.a.	3.00
<i>Hosta ventricosa</i>	C ₃	3	306.32±25.30	649.69±59.08	26.23±3.09	1.00
<i>Hosta ventricosa</i>	C ₃	3	264.40±20.49	662.35±28.98	27.14±2.25	1.00
<i>Manfreda scabra</i>	CAM	1	634.59±n.a.	1570.33±n.a.	5.79±n.a.	1.00
<i>Manfreda virginica</i>	CAM	3	1556.15±145.56	673.90±51.83	10.12±1.89	2.33
<i>Polianthes tuberosa</i>	C ₃	3	416.00±71.40	1133.18±46.86	8.22±1.23	1.00
<i>Polianthes tuberosa</i>	C ₃	2	419.40±24.59	1047.76±13.42	10.73±2.08	1.00
<i>Yucca aloifolia</i>	CAM	3	1326.41±131.62	1652.04±179.18	13.25±1.47	2.00
<i>Yucca angustissima</i>	C ₃	2	1394.68±821.49	595.05±20.42	5.11±1.36	3.50
<i>Yucca brevifolia</i>	C ₃	2	1121.51±368.25	591.13±46.98	2.67±0.49	1.50
CAM mean±SE		19	1294.14±76.87	1484.87±195.44	8.42±1.01	2.18±0.21
C ₃ mean±SE		24	509.26±115.31	897.21±47.38	18.99±2.27	1.46±0.18
T.test C ₃ vs. CAM			t=-5.95, df=35.41, p=0	t=-3.77, df=22.5, p=0.001	t=4.25, df=31.42, p=0.0002	t=-2.45, df= 32.97 p=0.02

¹number of images analyzed per plant sample

²no value could be calculated due to cross section imperfections (e.g., ripped section)

Figure 1 – Phylogenetic reconstruction of 68 taxa from the Agavoideae, using a) 272 nuclear genes in ASTRAL, b) a RAxML on a concatenated supermatrix of nuclear genes, and c) the chloroplast genome of 69,398bp. Branches with bootstraps less than 50 were collapsed.

Figure 2 – Ancestral state reconstruction of photosynthetic pathway in the Agavoideae. Taxa without carbon isotope values were removed, with the exception of representatives of the genera *Hosta*, *Echeandia*, and *Beschorneria*. The phylogeny is derived from nuclear supermatrix analysis, but unlike in Fig. 1b, nodes with bootstrap support less than fifty were not collapsed. Blue circles are C₃, yellow are CAM, with the probability of each state at each node indicated in pie graphs. Nodes with greater than 75BS support in supermatrix analysis are strongly bolded, while nodes with BS support of less than 75 but greater than 50 are a medium thickness.

Figure 3 – Principal coordinates analysis of the 19 BioClim climate variables for species in the Agavoideae. Points are coded by whether they are C₃ (blue) or CAM (yellow) based on isotopes assessed for an exemplar species.

Figure 4 – Cross section images from selected Agavoideae species. Phylogenetic tree to the left indicates relationships, and circles on tips indicate C₃ (blue) or CAM (yellow). For species that have high leaf heterogeneity in terms of thickness, both a central, thicker portion of the leaf as well as the thinner outer leaf are shown. Scale bar = 200µm.

Figure1

Figure2

Figure3

Figure4